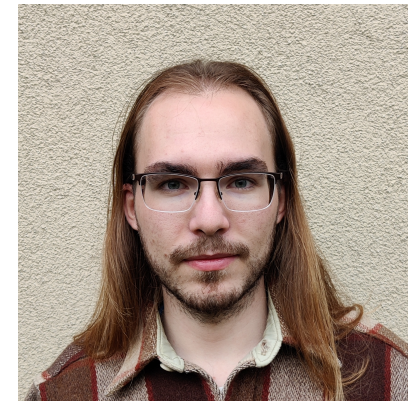


Jak nám jezdí vlaky?

David Koňářík · OpenAlt 2022

O mně

David Koňářík



Student MFF UK

Autor JrUtil, softwarového projektu na zpracovávání dat veřejné dopravy, a spousty žádostí o informace.

<https://dvdkon.gitlab.io>

dvdkon@konarici.cz

Téma

Zpracování dat o reálném provozu vlaků.

- Co mám jako cestující čekat?
- Kde dochází k problémům?
- Jak zlepšit vlakovou dopravu?
- + zajímavé “špeky”

Příklad dat

Vlak Os 10455 dne 2022-08-20

Zastávka	prav.příj.	příjezd	prav.odj.	odjezd
<i>P.-Velká Chuchle z</i>	13:44	13:43	13:44	13:44
<i>Odb Závodiště</i>	13:45	13:45	13:45	13:45
<i>Odb Barrandov</i>	13:47	13:48	13:47	13:48
<i>Pr.-Smíchov již.zhl.</i>	13:50	13:49	13:50	13:49
<i>Praha-Smíchov</i>	13:53	13:51	13:54	13:54
<i>Pr.-Smíchov sev.zhl.</i>	13:55	13:55	13:55	13:55
<i>Výh Praha-Vyšehrad</i>	13:57	13:57	13:57	13:57
<i>Praha hl.n.</i>	14:01	14:01		

Rozsah dat

Skoro kompletní data za zhruba poslední rok (1. 8. 2021–2022).

16 897 čísel vlaků, 2 718 257 jízd ~7 000 000 celkem minut zpoždění v koncové zastávce, ~900 000 minut náskoku, ~140 sekund průměrné zpoždění (včetně náskoků), 0 minut medián zpoždění.

Dálkové vlaky

1 135 čísel vlaků, 87 107 jízd ~1 1000 000 celkem minut
zpoždění v koncové zastávce, ~37 000 minut náskoku,
~740 sekund průměrné zpoždění (včetně náskoků), 5
minut medián zpoždění.

Nejzpožděnější vlaky

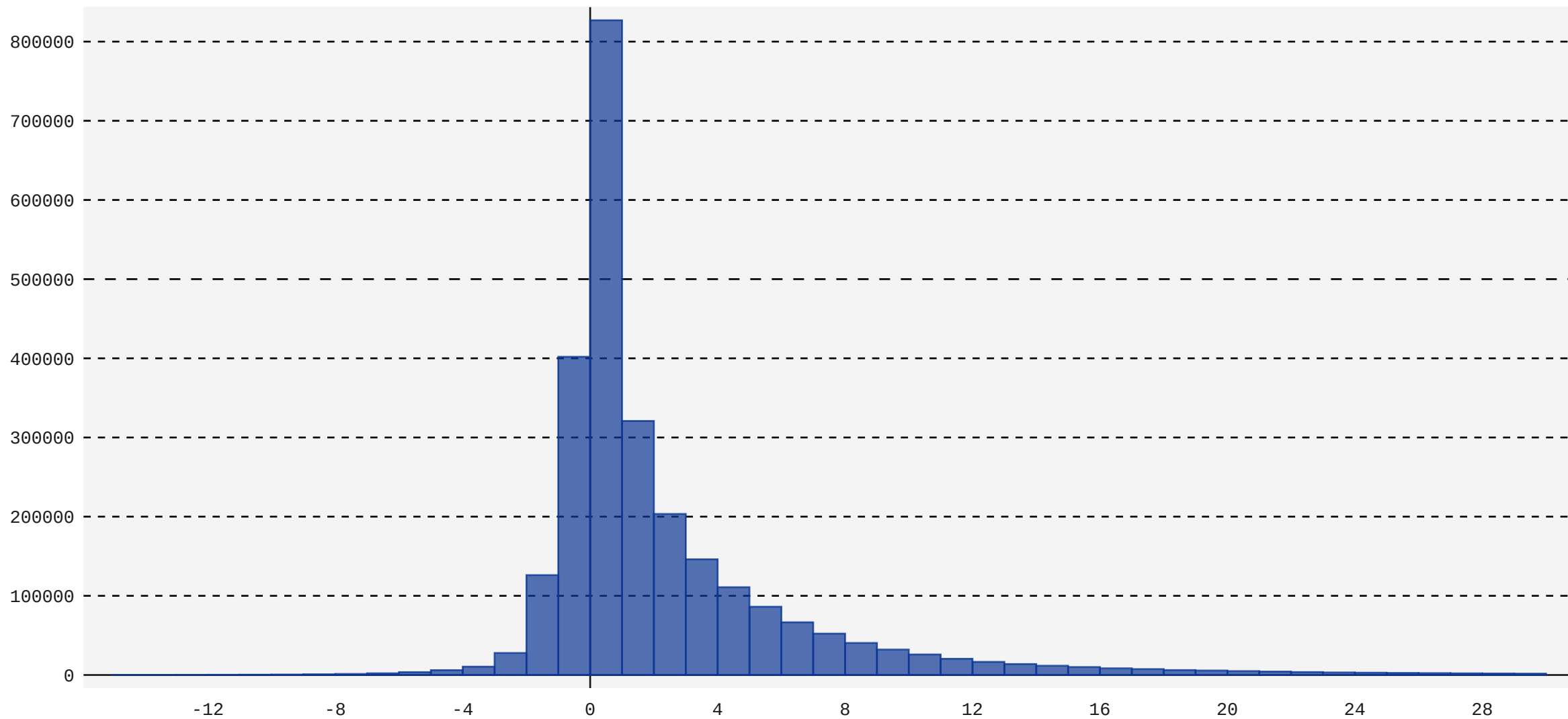
Vlak	Den	Zastávka	Zpoždění	Poznámka
Ex 11986	2022-02-28	Mosty u Jablunkova	19:11:00	Špatný JŘ?
Ex 43774	2022-02-28	Mosty u Jablunkova	19:11:00	Nedojel, špatný JŘ?
R 1122 RegioJet	2022-04-29	Bohumín os.n.	19:00:00	Jediný záznam
Os 18301	2022-06-10	Kralupy nad Vltavou	14:20:00	Jediný záznam
Os 7906	2022-02-17	Březnice	14:07:00	Nedojel
Ex 566 Západní expres	2022-02-17	Plzeň hl.n.os.n.	13:16:00	Nedojel
Os 5055	2022-03-01	Choceň	13:14:00	Jediný záznam
Os 21247	2022-01-08	Včelnička	12:56:00	Špatný JŘ?

Nejzpožděnější (rozumnější) vlaky

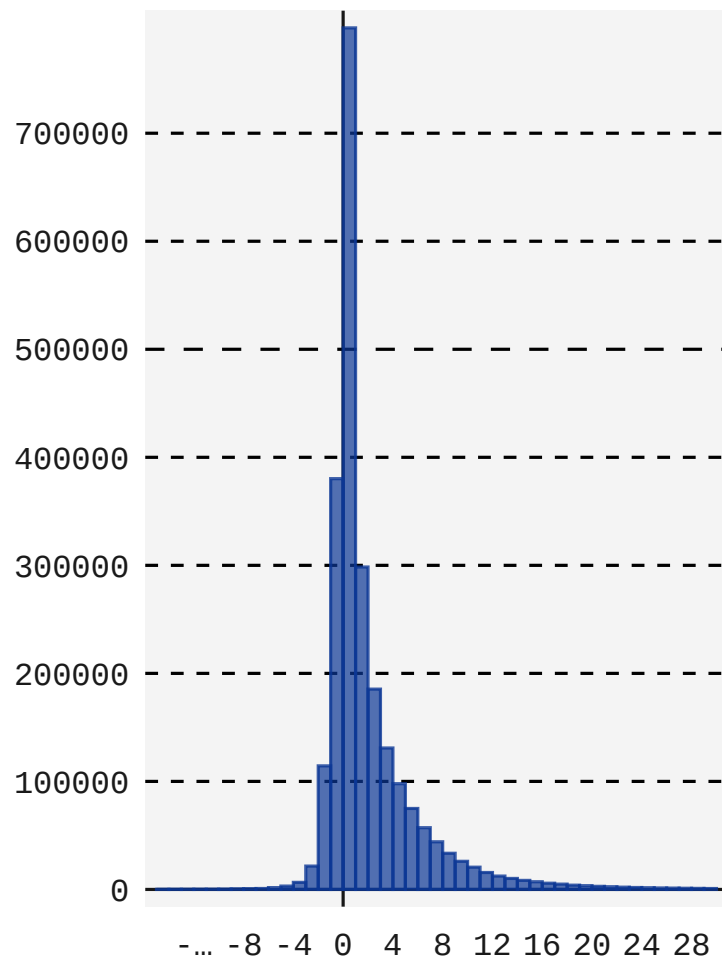
Vlaky, které dojely a měly na odjezdu z výchozí stanice max. 5h zpoždění.

Vlak	Den	Zastávka	Zpoždění
Os 13857	2021-09-13	Litovel	09:23:00
R 607 Krušnohor	2022-02-17	Odb Dubina	07:45:00
Os 96005	2021-10-20	Šakvice	07:39:00
SC 506 Pendolino	2022-02-17	Praha hl.n.	07:38:00
R 705 Vltava	2022-02-17	Č.Buděj. býv. st.7	07:10:00
Os 6058	2022-01-16	M.Boleslav m. předn.	07:00:00
Os 21241	2021-12-26	hi JH-Skrýchov JHMD	06:56:00

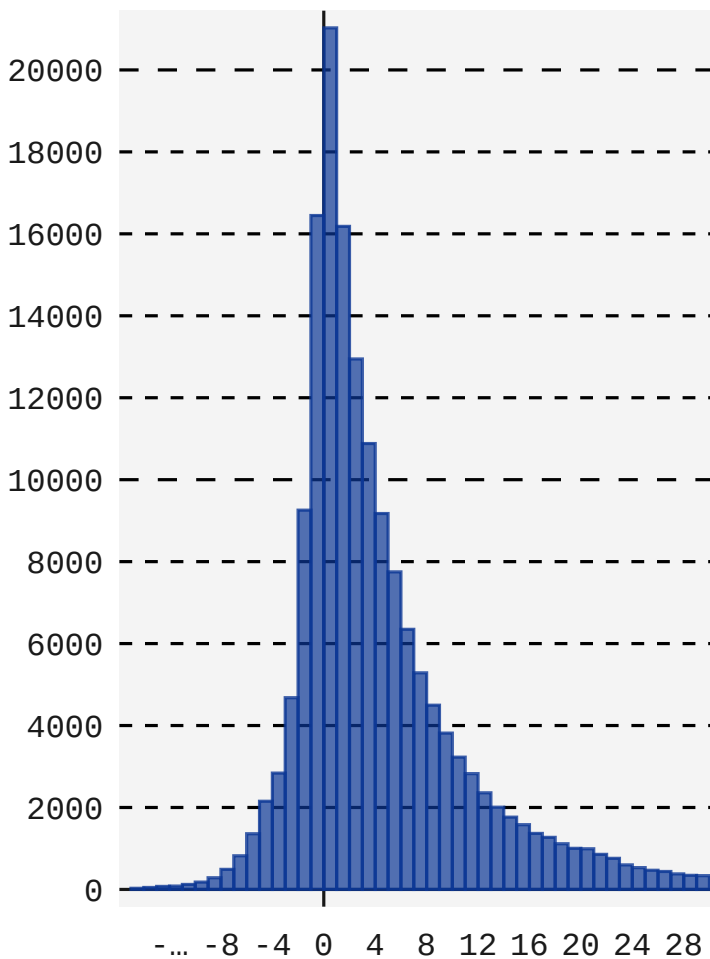
Histogram zpoždění



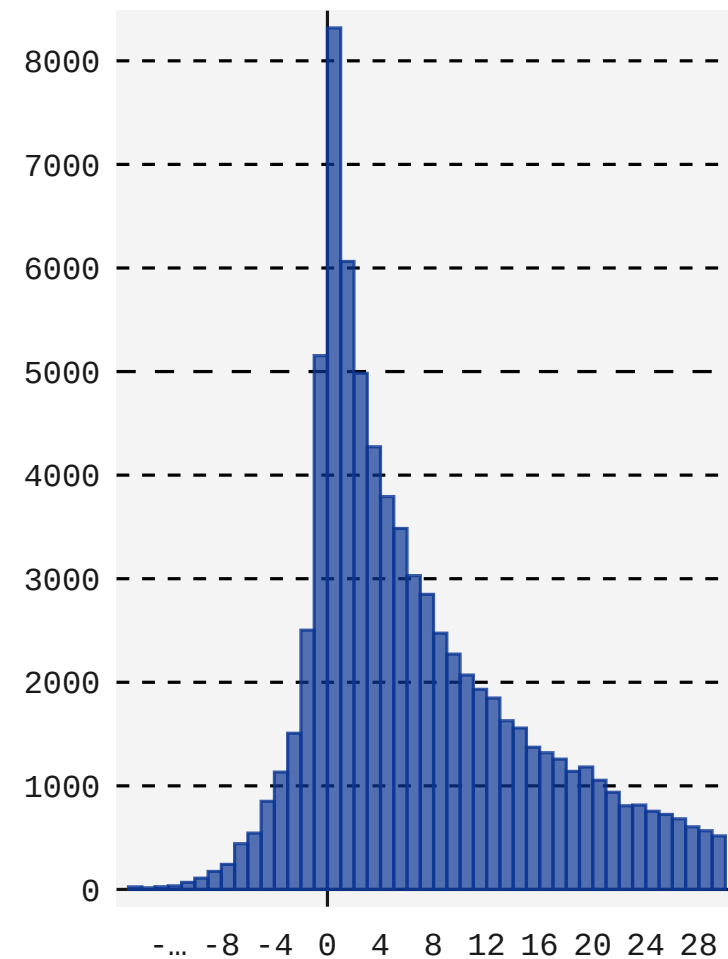
Histogram zpoždění



Osobní vlaky

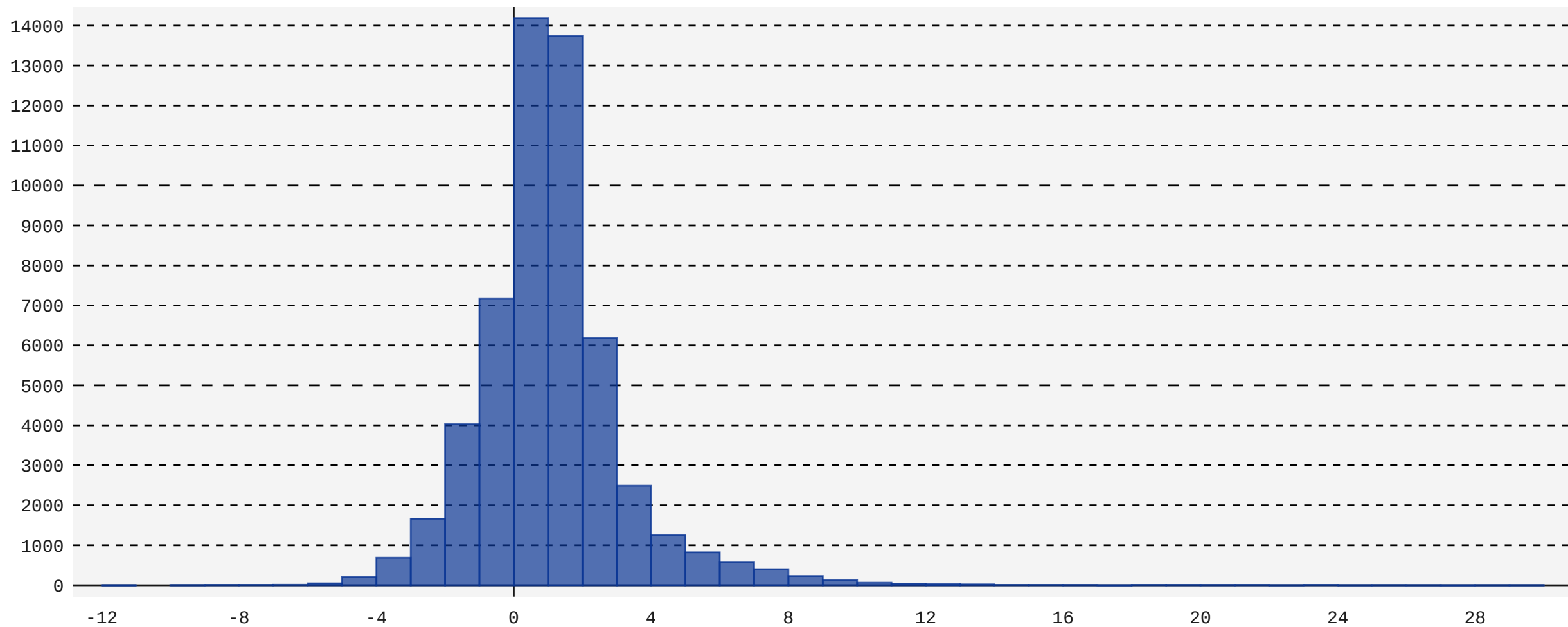


Vlaky R



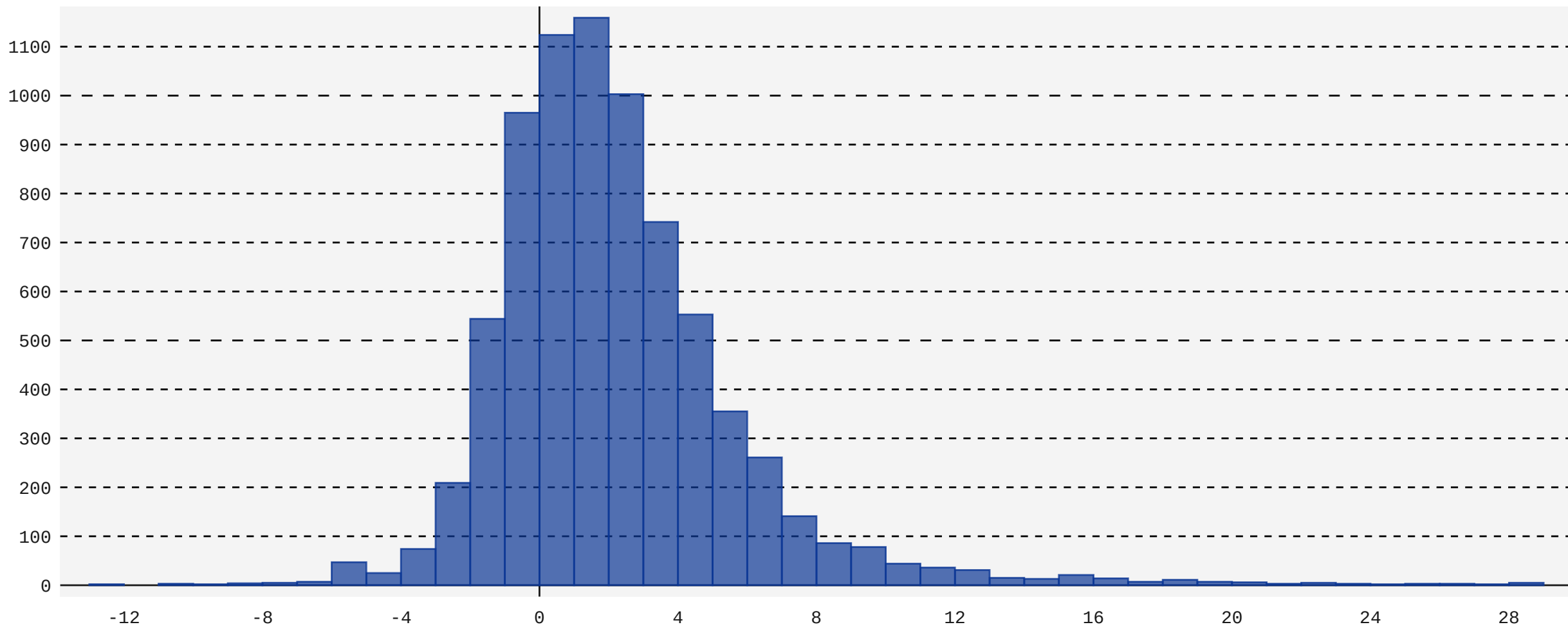
Dálkové vlaky

Histogram změny zpoždění



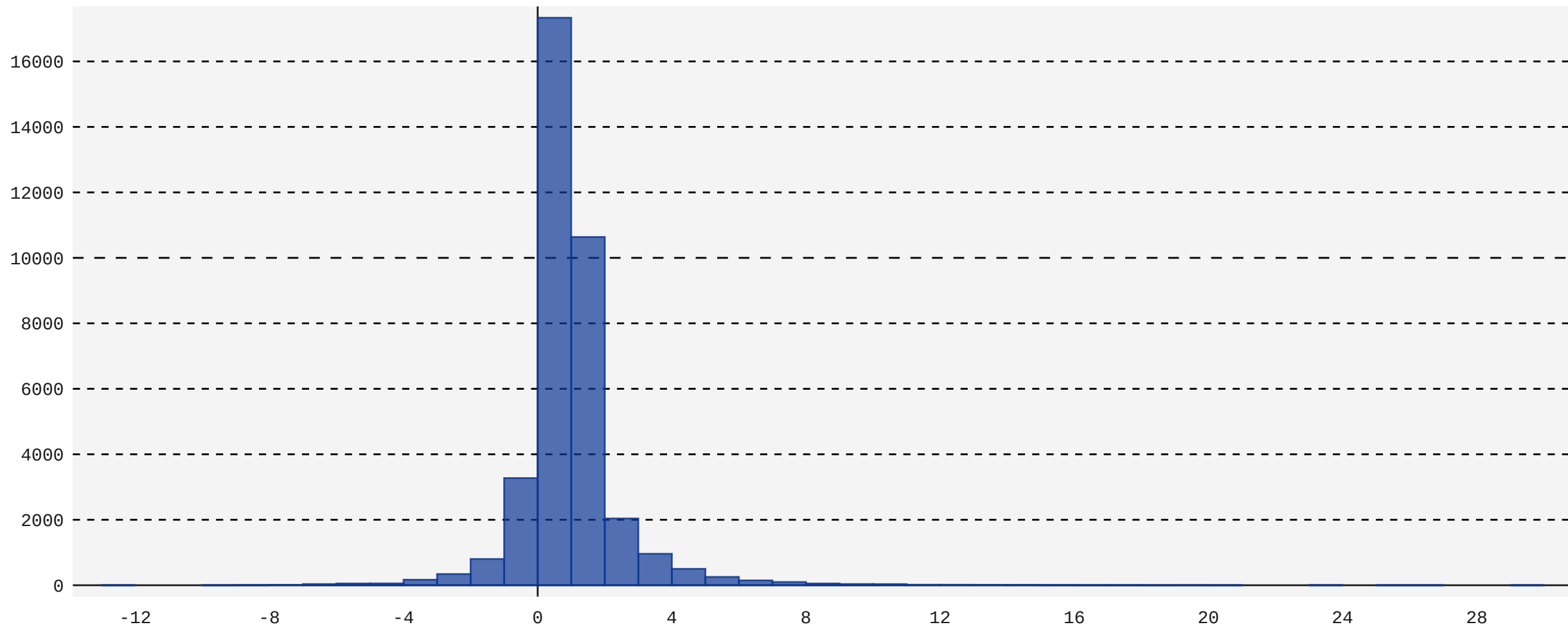
Brno hl.n. (28 000 : 26 000)

Histogram změny zpoždění



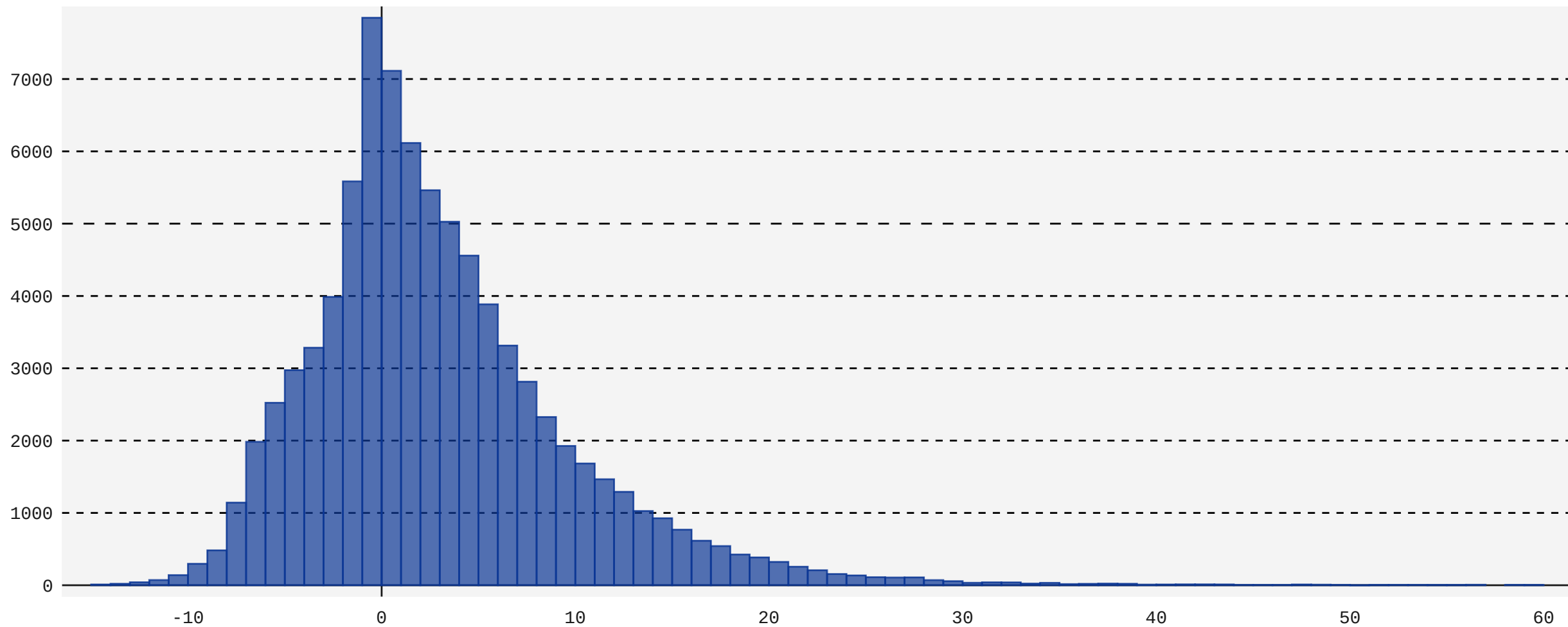
České Budějovice (3 000 : 4 600)

Histogram změny zpoždění



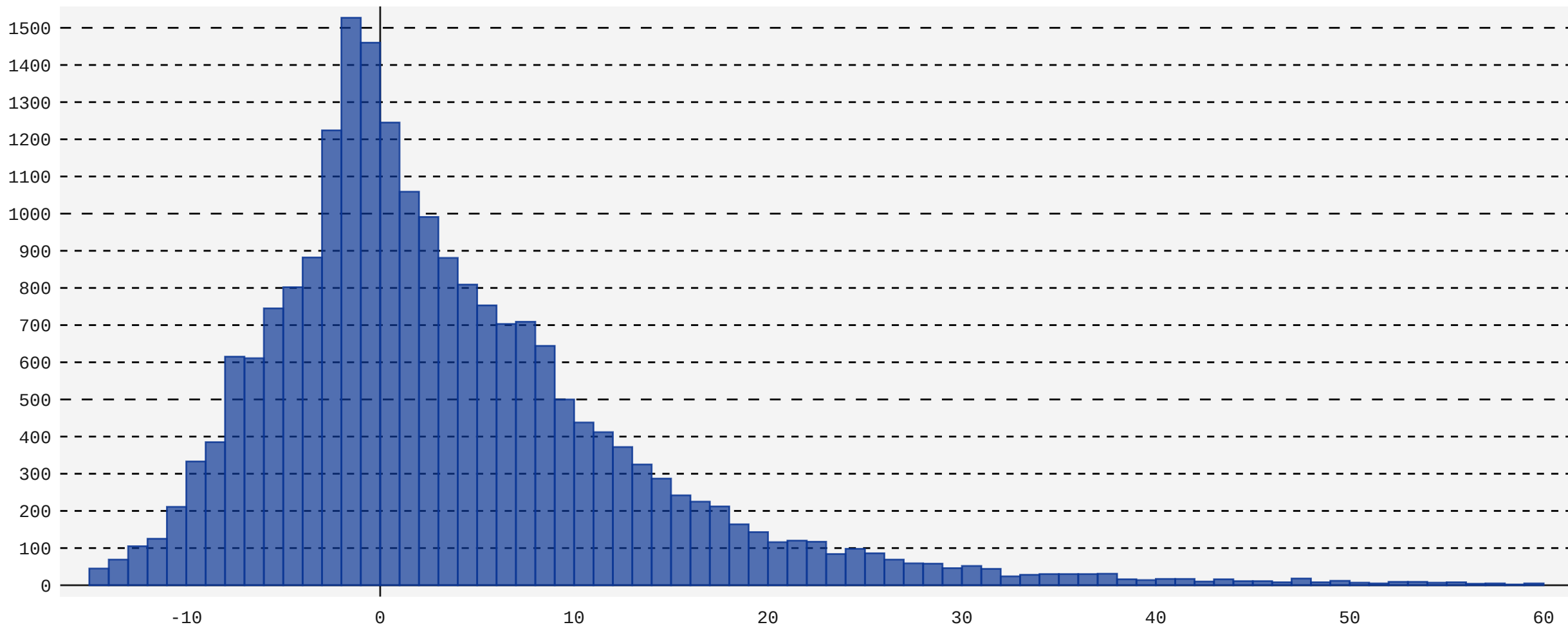
Hostivice (22 000 : 14 800)

Histogram změny zpoždění



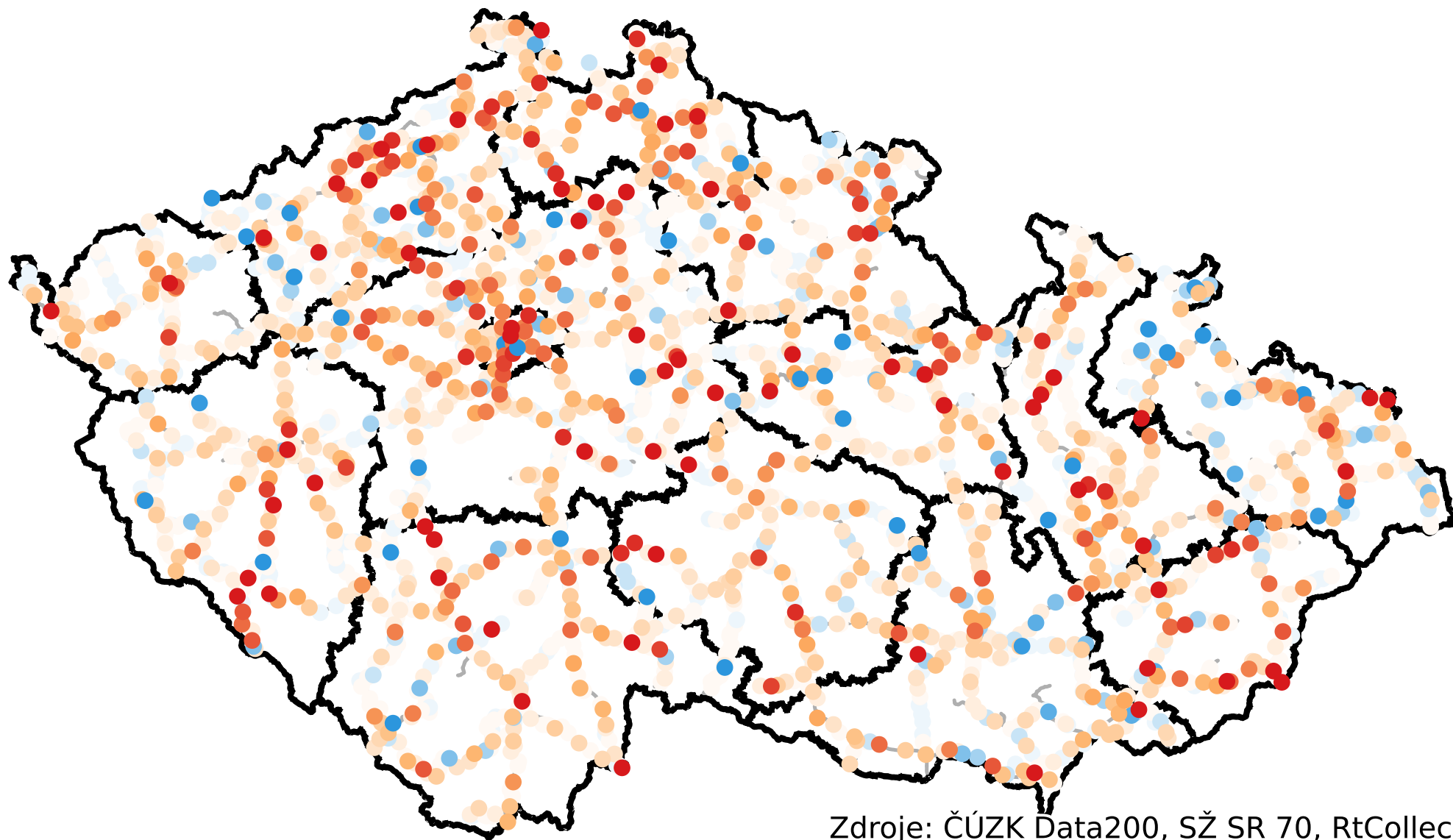
Úvaly → Kolín

Histogram změny zpoždění



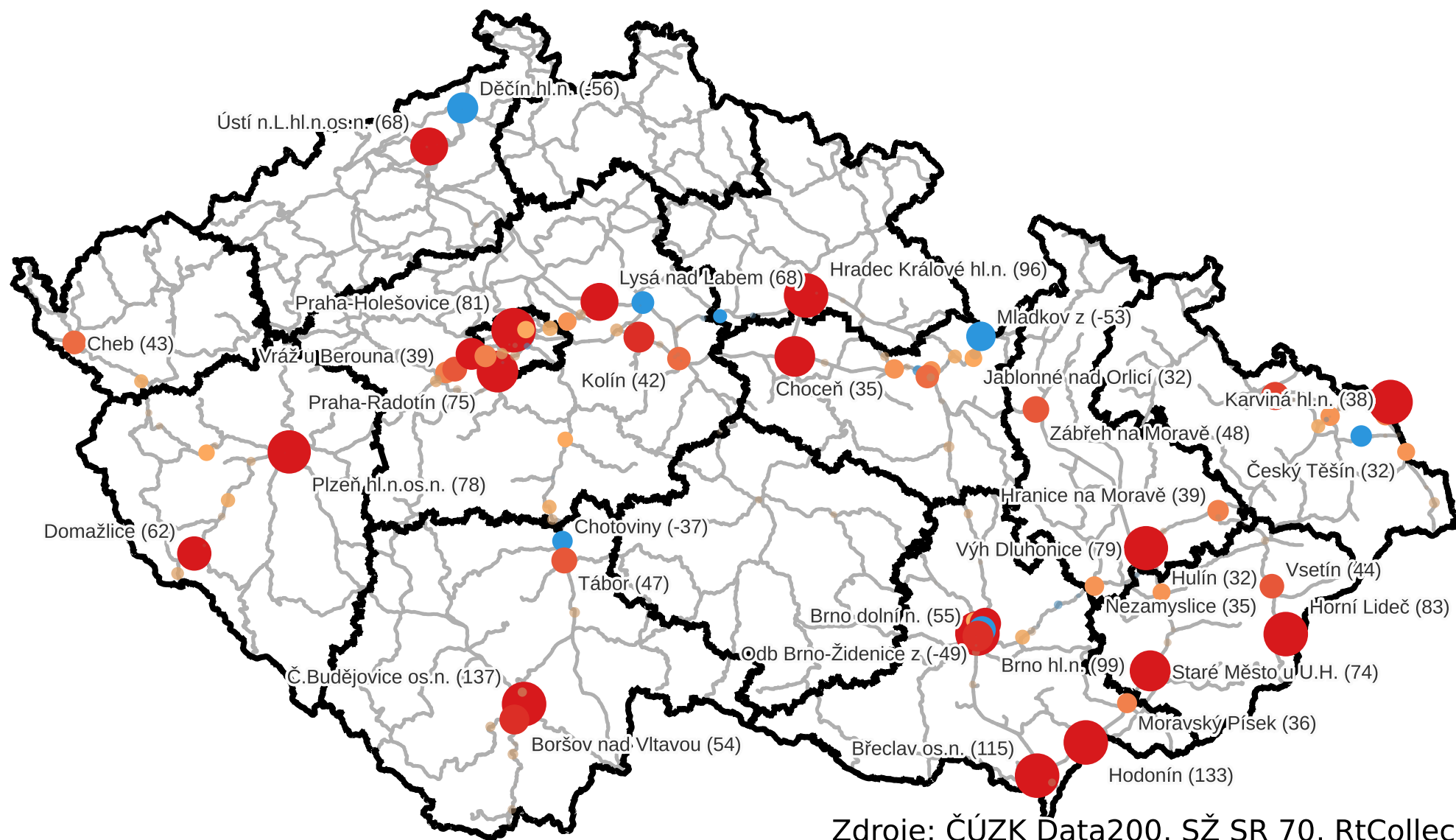
Praha hl.n. → Brno hl.n.

Změna zpoždění v zastávce

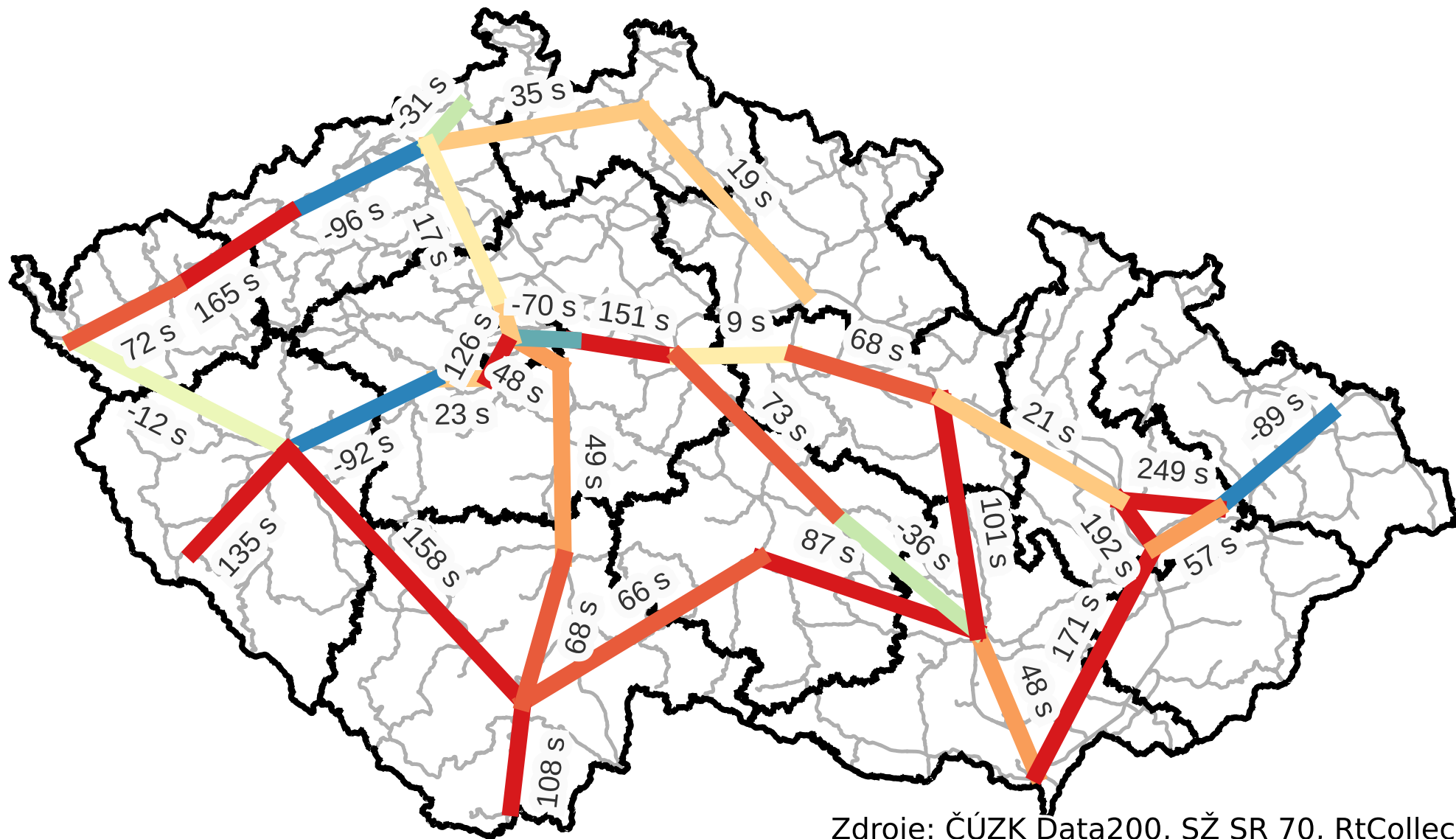


Zdroje: ČÚZK Data200, SŽ SR 70, RtCollect

Změna zpoždění (dálkové vlaky)



Změna zpoždění na trase



Zdroje: ČÚZK Data200, SŽ SR 70, RtCollect

Čištění dat

Čas od času opravdu divné vlaky, například:

- Chybné časy “reálného” odjezdu/příjezdu
- Chybné časy pravidelného odjezdu/příjezdu
- Neveřejné části trasy s notným náskokem

A taky chyby sbírání dat, nejčastěji špatně určené datum.

Vesměs hrubě promazáno.

Zdroj(e) dat

RtCollect, součást JrUtil, periodicky scrapuje všechny vlaky v GRAPPU.

CSV API RtCollect: <https://rt.jr.ggu.cz/api.html>

prosím opatrně, streamování nemusí fungovat, dotazy raději po měsících.

Většina vlaků v seznamu zůstane i desítky minut po příjezdu, přeshraniční ne.

Existuje i [SOAP rozhraní](#), ale...

Žadací dobrodružství

Podle zákona 106/1999 Sb. jsem SŽ žádal o několik věcí:

Žádost: “Aktuální polohy ... ve strojově čitelném formátu formou on-line přístupu”

Odpovědi:

- “... formou on-line přístupu poskytuje na základě smluvního vztahu.”
- “Informace ... je pouhou součástí informace o vlaku jako celku.”
- “Poskytnutí přístupu k většímu celku informací ... brání právní úprava zákona č. 181/2014 Sb.”

Žadací dobrodružství, kap. 2

Žádost: historické polohy vlaků

Odpověď:

- “Obsah žádosti překročil hranici mimořádně rozsáhlého vyhledávání informací.”
- “... stanovena předpokládaná časová náročnost shromáždění dostupných informací $305 \text{ hod} \times 325 \text{ Kč} = 99\,125 \text{ Kč}$.”

Žadací dobrodružství, spinoff

Vlakové JŘ jsou už [zveřejňovány](#) pořádně a úplně.

Včetně kolejí a nástupišť!

Linkové JŘ už ne: Nejprve nic, pak XLS, nyní osekané “skoro-JDF”.

CHAPS odmítá zákonnou povinnost, Ministerstvo buď odkazuje na ně, nebo popírá existenci úplných dat.

Není všem dnům konec: čekám na vyřízení stížnosti.

Kam dál

Data jsou k dispozici, těším se na vaše využití!

Po OpenAltu zveřejním kód na

<https://dvdkon.gitlab.io/openalt-2022-vlaky>

Fórum o datech veřejné dopravy:

<https://dadof.konarici.cz>

Všechny svoje další projekty/data budu oznamovat zde.

Poznámky na konec

V HTML verzi slidů jsou skryté poznámky:

<https://dvdkon.gitlab.io/openalt-2022-vlaky>